Final Report: Spatial Hearing, Attention and Informational Masking in Speech Identification.

INTRODUCTION

This report covers work supported by the above-referenced AFOSR award during the time period from June 1, 2005, through November 30, 2007. This was a collaborative effort between faculty and research staff at the Hearing Research Center, Boston University, and researchers at the Air Force Research Laboratory (AFRL/HE) at Wright-Patterson Air Force Base. The work consisted of both empirical and theoretical approaches primarily aimed at understanding the remarkable ability of humans to understand speech from one specific talker in the presence of competing talkers or other interfering sources of sound.

The final report draws upon a number of refereed publications and conference proceedings that are currently readily available in the scientific literature for detailed descriptions of the methods and findings from this research project. In instances where the work is not yet published more descriptive text is provided. These materials are organized and summarized according to the Specific Aims identified in the initial application with some additional related studies described as well.

SUMMARY OF ACCOMPLISHMENTS

Specific Aim 1: To examine the extent to which listeners are able to treat the two ears as independent sources of information that may be selectively attended to and whose inputs to the brain may be voluntarily controlled.

The human auditory system is commonly viewed as comprised of two distinct types of channels: the two ears form one type of channel and, within each ear, the tonotopic neural representation of frequency (i.e., auditory filters) forms the second type of channel. Auditory attention is often viewed as manifested in the ability to select the output of one or more channels and ignore the outputs from other channels. Informational masking (cf. Kidd et al., 2007) inherently reflects the inability of listeners in certain situations to ignore the irrelevant information in "masker" channels to the detriment of processing information in "target" channels. The work in this section addressed this issue directly through a series of speech identification experiments in which the speech was processed into narrow frequency bands so that it could be confined to specific auditory filters. In the article titled "The ability to listen with independent ears" Gallun et al. (2007a; see reference list below), examined a number of conditions under which it would be advantageous for a listener to ignore the input from one ear while processing the input to the other ear. Most modern models of binaural hearing explicitly incorporate monaural pathways that can be selectively attended to by the observer. Part of the evidence in support of these selectable monaural pathways comes from listening situations in which a performance advantage is found for the acoustically "better ear" resulting from head shadow. Gallun et al., however, reported several conditions in which listeners were unable to selectively attend to the better ear and appeared to be obliged to fuse similar information across the two ears. In particular, when the task was to identify speech processed into a set of narrow frequency bands and presented to one ear, the presentation of corresponding narrow bands of noise in the contralateral ear caused performance to suffer. This only occurred in a difficult listening situation in which the target speech had to be segregated from masking speech presented in the same ear. However, these findings mean the current models of the binaural processing of sounds are

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
28-02-08	Final	01/06/05-30/11/07

4. TITLE AND SUBTITLE

Spatial hearing, attention and informational masking in speech identification.

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-05-1-0337

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Gerald Kidd, Jr., Nathaniel Durlach, Douglas Brungart

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Trustees of Boston
University, 881 Commonwealth
Avenue, Boston, MA 02215

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFOSR
875 Randolph, Arlington, VA
22203

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-SR-AR-TR-08-0196

12. DISTRIBUTION / AVAILABILITY STATEMENT
DISTRIBUTION A: APPROVED FOR PUBLIC RELEASE

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The specific aims of this research grant included examining how and how well listeners could treat the two ears as independent channels, the role of *a priori* knowledge in multisource listening, and the evaluation of the Listener Max-Min observer models in multiple talker environments (Durlach et al., J. Acoust. Soc. Am., 2003). The general approach was empirical using humans subjects as observers in masked speech identification experiments when multiple sources of sound were present and, usually, were spatially distributed. As a general statement, we feel that considerable progress was made toward each of these aims. A list of publications and presentations at scientific meetings is appended. The work that was accomplished with AFOSR support has considerable scientific significance and possible relevance for practical applications. Furthermore, it has stimulated new lines of research in auditory attention and masked conditions where higher-level cognitive processes are key.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*

inaccurate and must be modified to incorporate stimulus-dependent segregation and task-dependent processing resource limitations. This work is leading us toward a model of auditory channel selection in which both bottom-up grouping principles and top-down attentional focus are primary (and sometimes competing) components.

Specific Aim 2: To examine how *a priori* knowledge about the characteristics of sound sources, and in particular their locations and frequency content, can lead to significant improvements in auditory performance in complex multisource listening situations.

In one published study (Kidd et al., 2005), we examined how uncertainty about the location of a target talker affected speech identification in a multitalker listening environment. The observers were positioned in a sound field with loudspeakers at three different locations. Three different sentences were presented on each trial and the task of the listener was to repeat back the key words of the target sentence which was identified by a specific callsign. The main parameter that was varied in the study was the degree of uncertainty about which of the three locations presented the target. When the target location was completely certain, speech identification performance was exceptionally good with scores in all conditions greater than 90% correct. Performance declined monotonically as the uncertainty about target location increased. This study demonstrated that *a priori* knowledge about the characteristics of a target talker - in this case talker location - can have very significant effects on the ability to select and attend to a specific source embedded in competing sources. It should be pointed out that this large effect of *a priori* knowledge is only observable in complex and uncertain listening environments. It is for this reason that such contextual effects have often been considered fairly minor factors in auditory tasks. This work proves otherwise.

Specific Aim 3: To evaluate the theoretical constructs of acceptance vs. rejection filters in auditory attention as they apply to speech recognition in multisource environments.

The notion of filtering in the spatial dimension - analogous to the well-known filtering in the frequency dimension - has been raised by several past investigators although until recently the evidence in support of this idea was not compelling (cf. Scharf, B., 1998, "Auditory attention: The psychoacoustical approach," in *Attention*, edited by H. Pashler, Hove, East Sussex: Psychology Press Ltd., pp. 75-117). Two studies addressing this issue were completed during the period of time covered by this final report. First, Marrone et al. (2008; conditionally accepted for publication) found strong evidence for auditory spatial filters that appeared to be related to the focus of attention in highly complex and uncertain listening situations. In that study, speech identification performance was compared in situations in which a target talker and two masking talkers were colocated and when the two masker talkers were spatially separated symmetrically from the target. Because of the high degree of informational masking present in this listening situation, spatial separation of sources provided a strong cue for segregating and focusing attention on the target. This effect was related to the degree of spatial separation of target and masker such that a pattern of release from masking was observed that showed a clear and significant tuned response. Marrone et al. fit filter functions to the data and concluded that the bandwidth of these spatial filters was quite narrow; for most subjects it was less than $\pm 10°$. When the masker talkers were replaced by noise - producing little informational masking but large amounts of energetic masking - very little spatial tuning was observed. This result was

interpreted as indicating that spatial filtering is largely a higher-level process (unlike the initial cochlear filtering in the frequency domain) that is most important in very complex and uncertain listening environments.

Although the study above provided strong evidence for spatial filtering, it did not distinguish between the two observer models proposed by Durlach et al. (2003) termed "Listener Max" and "Listener Min." Under certain conditions, either model could account for the tuned pattern of responses reported by Marrone et al. (2008). Recently, though, we have used a new approach to studying and contrasting these two models (Kidd et al., 2008; under review). Using a new modification of a procedure originally developed by Broadbent ("Failures of attention in selective listening," J. Exp. Psychol., 44, 428-433, 1952) target and masker speech streams were presented in an alternating word format. Thus, the target comprised the odd-numbered words in the sequence while the masker comprised the even-numbered words in the sequence. A variety of acoustical and syntactic "linkages" were used to bind either the target or masker words together. These linkage variables were very effective in overcoming the informational masking caused by the presence of the masker words, but only when applied to the target words. Thus, for example, holding the apparent location (determined by a fixed ITD) of the target constant throughout a trial improved performance considerably relative to the situation in which target location varied randomly. However, the same manipulation when applied to the masker yielded no improvement in performance. The interpretation of this result is that the Listener Max model in which the observer applies the available processing resources to enhance the representation of the target, provided a better explanation of the findings than did the Listener Min model in which the available processing resources are devoted to nulling or minimizing the masker. Although this conclusion seemed warranted based on the results, it should be mentioned that it is quite possible that a Listener Min strategy is adopted by observers in other tasks and further work is needed to understand how and when listeners employ one strategy versus the other.

Related Studies

Two additional articles describing work supported by this AFOSR award should be mentioned. Both are related to the specific aims of this work but do not fit as directly under any single aim as the studies above.

Gallun et al. (2007b) examined the costs associated with dividing attention between two sources and distinguished them from the costs of selectively attending to one source in the presence of a second unwanted source. Their study used the same type of processed speech described above and presented two different sentences to their observers on every trial with one sentence presented to one ear and the other sentence presented to the opposite ear. In selective listening conditions, the observer was instructed simply to detect or identify the speech in one ear while ignoring the opposite ear. In divided listening conditions, the observer had to monitor both ears in detection or identification tasks. Predictably, performance in the divided listening task was poorer than in the selective listening task although there was a cost of having an irrelevant distracting speech stimulus even in the selective listening condition. However, in the divided listening task the costs were much greater when the listener had to monitor both ears for speech identification than when the listener only had to identify the speech in one ear and detect the presence of speech in the opposite ear. Gallun et al. speculated that the costs of dividing attention is related to the extent to which the two tasks require the same or different pools of processing

resources. So, when two identification tasks were required the observer was tapping the same pool of resources whereas when the observer was making one identification judgment and one detection judgment different pools of resources were tapped.

Best et al. (2007) also examined both selective and divided attention in an auditory identification task. In their experiments, the observer was required to report key words from one talker in the presence of a second talker (selective listening) or report the key words from both talkers (divided listening). They main variables they manipulated were the relative levels of the two sources, the spatial separation of the sources, and the presence/level of a Gaussian noise added to the speech. They found that spatial separation of sources improved performance not only in the selective listening task but also in the divided listening task. This result was not consistent with the idea that a single attentional spotlight alternated between sources because the opposite pattern of results would be predicted. Instead, the ability to solve the divided listening task appears to depend on source resolution and the strength of segregation of the two sources. Furthermore, adding noise to the speech had a significantly greater negative effect on performance in the divided listening task for the stimulus that the observer reported second compared to the stimulus that the observer reported first. This result was interpreted as evidence for the noise adding to the decay of the sensory trace of the second-reported stimulus that must be held in a memory store while the first stimulus is reported.

LIST OF PRIMARY CONTRIBUTORS

At Boston University:

Gerald Kidd, Jr., Ph.D., Co-Principal Investigator
Nathaniel I. Durlach, M.S., Co-Principal Investigator
Christine R. Mason, M.S., Co-Investigator
Virginia Best, Ph.D., Co-Investigator
Antje Ihlefeld, Ph.D., Co-Investigator
Nicole Marrone, Ph.D., Graduate Research Assistant

At WPAFB (note: the research activities of these investigators are described in separate reports)

Douglas Brungart, Ph.D., Co-Principal Investigator
Brian Simpson, Ph.D., Co-Investigator
Nandini Iyer, Ph.D., Co-Investigator

LIST OF PUBLICATIONS

Book Chapter

Kidd, G. Jr., Mason, C.R., Richards, V. M., Gallun, F. J. and Durlach, N.I. "Informational masking," in *Auditory Perception of Sound Sources*, Yost, W. A., Popper, A. N., and Fay, R. R. (eds). (2007). Springer Science+Business Media, LLC, New York, 143-190

Published Articles

Kidd, G. Jr., Arbogast, T.L., Mason, C.R. and Gallun, F.J. (2005) "The advantage of knowing where to listen," J. Acoust. Soc. Am., 118, 3804-3815

Colburn, H.S., Shinn-Cunningham, B.A., Kidd, G. Jr. and Durlach, N.I. (2006) "The perceptual consequences of binaural hearing," Int. J. Audiology, 45, 34-44

Gallun, F.J., Mason, C.R. and Kidd, G. Jr. (2007a) "The ability to listen with independent ears," J. Acoust. Soc. Am., 122, 2814-2825

Gallun, F.J., Mason, C.R. and Kidd, G. Jr. (2007b) “Task-dependent costs in processing two simultaneous auditory stimuli” Percept. Psychophys., 69, 757-771

Best, V., Ihlefeld, A., Mason, C.R., Kidd, G. Jr. and Shinn-Cunningham, B.G. (2007) "Divided listening in auditory displays," Proceedings of the International Congress on Acoustics, Madrid, Spain

In Press

Marrone, N.L., Mason, C.R. and Kidd, G. Jr. (2007) "Listening in a multisource environment with and without hearing aids," Proceedings of the International Symposium on Auditory and Audiological Research, in press

Marrone, N.L., Mason, C.R. and Kidd, G. Jr. (2008) "Tuning in the spatial dimension: Evidence from a masked speech identification task" J. Acoust. Soc. Am. (conditionally accepted for publication)

Gallun, F.J., Durlach, N.I., Colburn, H.S., Shinn-Cunningham, B.G., Best, V., Mason, C.R. and Kidd, G. Jr. (2008) "The extent to which a position-based explanation accounts for binaural release from informational masking" J. Acoust. Soc. Am. (conditionally accepted for publication)

Under Review

Kidd, G. Jr., Best, V. and Mason, C.R. "Listening to every other word: Examining the strength of linkage variables in forming streams of speech"